Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes?

Heiskanen, Janne

2019-09

Heiskanen , J , Adhikari , H , Piiroinen , R , Packalen , P & Pellikka , P K E 2019 , ' Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes? ' , International Journal of Applied Earth Observation and Geoinformation , vol. 81 , pp. 176-185 . https://doi.org/10.1016/j.jag.2019.0

http://hdl.handle.net/10138/302808
https://doi.org/10.1016/j.jag.2019.05.017

Int J Appl Earth Obs Geoinformation 81 (2019) 176–185

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes?

Janne Heiskanen[a,b,*], Hari Adhikari[a,b], Rami Piiroinen[a,b], Petteri Packalen[c], Petri K.E. Pellikka[a,b]

[a] Department of Geosciences and Geography, University of Helsinki, P.O. Box 68, FI-00014, Finland
[b] Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki, Finland
[c] School of Forest Sciences, University of Eastern Finland, Yliopistokatu 7, FI-80100, Joensuu, Finland

ARTICLE INFO

Keywords:
Africa
Airborne laser scanning
AisaEAGLE
Carbon
Imaging spectroscopy
Landsat
LiDAR

ABSTRACT

Airborne laser scanning (ALS) is considered as the most accurate remote sensing data for the predictive modelling of AGB. However, tropical landscapes experiencing land use changes are typically heterogeneous mosaics of various land cover types with high tree species richness and trees outside forests, making them challenging environments even for ALS. Therefore, combining ALS data with other remote sensing data, or stratification by land cover type could be particularly beneficial in terms of modelling accuracy in such landscapes. Our objective was to test if spectral-temporal metrics from the Landsat time series (LTS), simultaneously acquired hyperspectral (HS) data, or stratification to the forest and non-forest classes improves accuracy of the AGB modelling across an Afromontane landscape in Kenya. The combination of ALS and HS data improved the cross-validated RMSE from 51.5 Mg ha^{-1} (42.7%) to 47.7 Mg ha^{-1} (39.5%) in comparison to the use of ALS data only. Furthermore, the combination of ALS data with LTS and HS data improved accuracies of the models for the forest and non-forest classes, and the overall best results were achieved when using ALS and HS data with stratification (RMSE 40.0 Mg ha^{-1}, 33.1%). We conclude that ALS data alone provides robust models for AGB mapping across tropical mosaic landscapes, even without stratification. However, ALS and HS data together, and additional forest classification for stratification, can improve modelling accuracy considerably in similar, tree species rich areas.

1. Introduction

Large amounts of carbon are stored in the aboveground biomass (AGB) in tropical forests, and change in the tropical forest cover is a major source of carbon emissions (Houghton et al., 2009; Ciais et al., 2013). Although most of the AGB resides in the forests, the importance of the trees outside forests has been increasingly underlined (Schnell et al., 2015; Sloan and Sayer, 2015; Vanderhaegen et al., 2015). With the agricultural expansion and intensification, tropical landscapes become more affected by human activities, and turn into the mosaics of agricultural land, plantations and secondary, logged and fragmented forests (Mertz et al., 2012; Laurance et al., 2014). Trees are also common in the tropical agro-ecological systems, and agroforestry and trees on croplands make major contributions to the global and regional carbon budgets (Albrecht and Kandji, 2003; Zomer et al., 2016). Therefore, in order to better understand AGB distributions and climate change mitigation possibilities within the landscapes, it is important to depict AGB variations also outside forests.

Remote sensing is required for mapping AGB at most of the scales. The current pan-tropical and continental-scale maps have relatively coarse resolution (Saatchi et al., 2011; Baccini et al., 2012; Avitabile et al., 2016), large uncertainties in the pixel-level (Baccini et al., 2017), or are restricted to the savannas with relatively low AGB (Bouvet et al., 2018). Hence, these data are not yet sufficient to characterize AGB variations in the landscapes where small forest patches can have high AGB. In general, AGB is predicted most accurately by using light detection and ranging (LiDAR) systems, in particular, the airborne laser scanning (ALS) (Zolkos et al., 2013). The feasibility of ALS has been demonstrated for different vegetation types in Africa, including forests with high AGB (Asner et al., 2012; Vaglio et al., 2014, 2016; Hansen et al., 2015; Fatoyinbo et al., 2018) and savannahs with lower AGB (Mauya et al., 2015; Næsset et al., 2016; Tesfamichael and Beech, 2016; Egberth et al., 2017). However, rather few studies have so far examined agriculture-forest mosaics including both high AGB forests and almost

* Corresponding author at: Department of Geosciences and Geography, University of Helsinki, P.O. Box 68, FI-00014, Finland.
E-mail address: janne.heiskanen@helsinki.fi (J. Heiskanen).

https://doi.org/10.1016/j.jag.2019.05.017
Received 31 March 2019; Received in revised form 16 May 2019; Accepted 20 May 2019

treeless vegetation types in Africa (Asner et al., 2012; Adhikari et al., 2017; Graves et al., 2018).

One possibility to improve the accuracy of the predictive models for AGB is to combine ALS and optical remote sensing data (Zolkos et al., 2013). Several studies have used Landsat data for AGB mapping in Africa, including the latest Landsat 8 OLI imagery (Dube and Mutanga, 2015; Karlson et al., 2015; Gizachew et al., 2016; Egberth et al., 2017; Heiskanen et al., 2017). Egberth et al. (2017) combined ALS and Landsat 8 OLI imagery in the Miombo woodlands in Tanzania, which improved accuracy slightly in comparison to the ALS models. However, a greater synergy of ALS and Landsat data was found by Phua et al. (2017) for the tropical forest in Malaysia. Furthermore, with the free of cost Landsat data, it has become more feasible to utilize a time series of imagery instead of single scenes. The spectral-temporal metrics based on Landsat time series (LTS) have potential to improve the accuracy of the AGB prediction models through complementary information on the land surface phenology (Zhu and Liu, 2015). The potential of such metrics has been demonstrated for mapping tree height, an important factor in AGB, across Africa (Hansen et al., 2016).

Hyperspectral (HS) remote sensing (imaging spectroscopy) has been shown to be highly useful for the classification of tree species in forests (Fassnacht et al., 2016) and agro-ecosystems (Piiroinen et al., 2015; Graves et al., 2016; Piiroinen et al., 2017). Although HS data have not usually performed well in AGB modelling alone (Clark et al., 2011; Zolkos et al., 2013), the integration of HS data to AGB prediction models could improve modelling accuracy as it includes complementary information on tree species composition. For example, Vaglio et al. (2014) found out that the combination of ALS and HS data outperformed ALS data in the tropical forest in Sierra Leone. Therefore, there could be synergies in some cases, for example, when tree species richness is particularly high. However, the integration of both ALS data with LTS and HS data remain poorly explored in the forest-agricultural land use mosaics in tropical Africa.

When modelling AGB across the mosaic landscapes, it is also relevant to consider stratification by vegetation type or land use class (Mauya et al., 2015; Latifi et al., 2015). AGB, canopy structure, and tree species composition can vary considerably between the strata with an effect on the characteristics of ALS point clouds and the most suitable predictor variables. Earlier, some studies have shown that stratification can improve ALS models for forest AGB (Clark et al., 2011) while others have found only marginal benefits (Latifi et al., 2015). Mauya et al. (2015) showed that AGB predictions were improved by the post-stratification in the Miombo savannah in Tanzania. Therefore, stratification should be also studied for other landscape types in Africa.

Our main objective was to examine if ALS data complemented with LTS or HS data, or forest - non-forest classification, improves the accuracy AGB predictions across an Afromontane landscape in Kenya. More specifically, we studied how accurately AGB can be modelled using ALS, LTS and HS variables separately and together, and modelled AGB separately for the forest and non-forest classes using different data combinations. Finally, we compared non-stratified models (model fit without considering forest classification) and stratified models (predictions for the forests and non-forest classes combined) at the landscape-level.

2. Material and methods

2.1. Study area

We conducted this study in the Taita Hills, in southeastern Kenya (Fig. 1). The data were collected from a 10 km × 10 km study area in the middle of the hills. The area is topographically variable and altitude range from approximately 1000 m a.s.l. to the highest peak of the Taita Hills, Vuria (2208 m a.s.l.). The area has a bimodal rainfall pattern with rainy seasons from October to December and from late March to June. The amount of rainfall varies according to the topography but it is greater than 1200 mm per year in the hills (Pellikka et al., 2013).

The Taita Hills belong to the Eastern Arc Mountains with high conservation value (Platts et al., 2011). The Afromontane forests are particularly valuable in terms of biodiversity and other ecosystem services. However, the area is currently densely populated and have undergone large scale deforestation. The remaining forest patches are very fragmented and located near the mountaintops (Pellikka et al., 2009, 2013). Some of the most common native species include *Tabernaemontana stapfiana*, *Macaranga capensis*, *Oxyanthus speciosus*, *Phoenix reclinata* and *Celtis africana*. Otherwise, the landscape consists of smallholder agriculture (mainly maize fields), the plantations of exotic trees (*Eucalyptus* spp., *Pinus patula*, *Cupressus lusitanica* and *Acacia mearnsii*), bushland and settlements (Fig. 2). Agroforestry and trees outside forests (e.g. *Grevillea robusta*) are common in the study area.

2.2. Field measurements and aboveground biomass calculation

In order to estimate reference AGB for the modelling, we surveyed 131 circular 0.1 ha plots in January and February in 2013 and 2014. The sampling scheme (Fig. 1) varied between the two field campaigns. In 2013, we sampled ten clusters of ten plots using geographically stratified random sampling (Vågen et al., 2013). In 2014, we set up additional plots to forests based on ALS data and HS imagery. The plot centers were located by a Trimble GeoXH GNSS receiver with differential correction (see Pellikka et al., 2018 for details).

Within the plot radius (17.84 m), we measured stem diameter at breast height (*D*) and identified tree species for all living stems having $D > 10$ cm. The stem diameter was measured using a diameter measurement tape. We measured tree height (*H*) for most of the trees outside forests by a hypsometer (Suunto). In 2014, *H* was measured for at least three sample trees (minimum, maximum and median *D*) in the forest plots using a laser range finder (Laser Technology TruPulse 360). However, ten forest plots surveyed in 2013 lacked *H* measurements and we estimated maximum *H* from the ALS-based canopy height model (CHM) and assumed that the tree with the largest *D* was also the tallest tree in the plot. Based on the *H* measurements, we imputed *H* for all the tallied trees by using the two-parameter Curtis's function (Curtis, 1967), and non-linear mixed effect modelling and plot as random effects (Valbuena et al., 2016). The models were fit using 'nlme' package in R (Pinheiro et al., 2015) with the root mean square error of 3.8 m (30.5%). For palms, we always measured *H* because *D* is not good predictor of *AGB* (Brown, 1997).

As the main allometric equation for *AGB* estimation, we used the pan-tropical model of Chave et al. (2014) based on *D*, *H* and wood density (ρ). The model includes also data from the Eastern Arc Mountains (Marshall et al., 2012). We searched ρ from Zanne et al. (2009) and ICRAF (2015). In total, we recorded 95 species in the plots and found species or family level ρ value for 84 species (96.1% of all the stems). For the remaining species, we employed a landscape-level mean value (0.562 g cm^{-3}). The data used in the Chave et al. (2014) model does not include tropical plantations. Therefore, we used additional species-wise equations for *A. mearnsii* and *Eucalyptus* spp. (Paul et al., 2013). For pines (*Pinus* spp.), we applied a stem volume equation from Henry et al. (2011: equation 474) and default biomass expansion factor for tropical pines (IPCC, 2003). The palm equation based on *H* (Brown, 1997) was used for *P. reclinata*. See Pellikka et al. (2018) for more details on the allometric models. Lastly, we summed all the tree-wise *AGB* values for each plot and calculated AGB per hectare according to the plot area. We did not make slope correction in the field, and hence, corrected the plot area based on the slope derived from the digital elevation model (DEM).

Table 1 presents summary statistics for the plots. Nine plots in 2013 data were treeless, and one plot was removed as we considered *D* measurements of two very large trees unreliable because of the abundant lianas. Furthermore, four plots lacked valid HS data. This resulted in 117 plots for modelling.

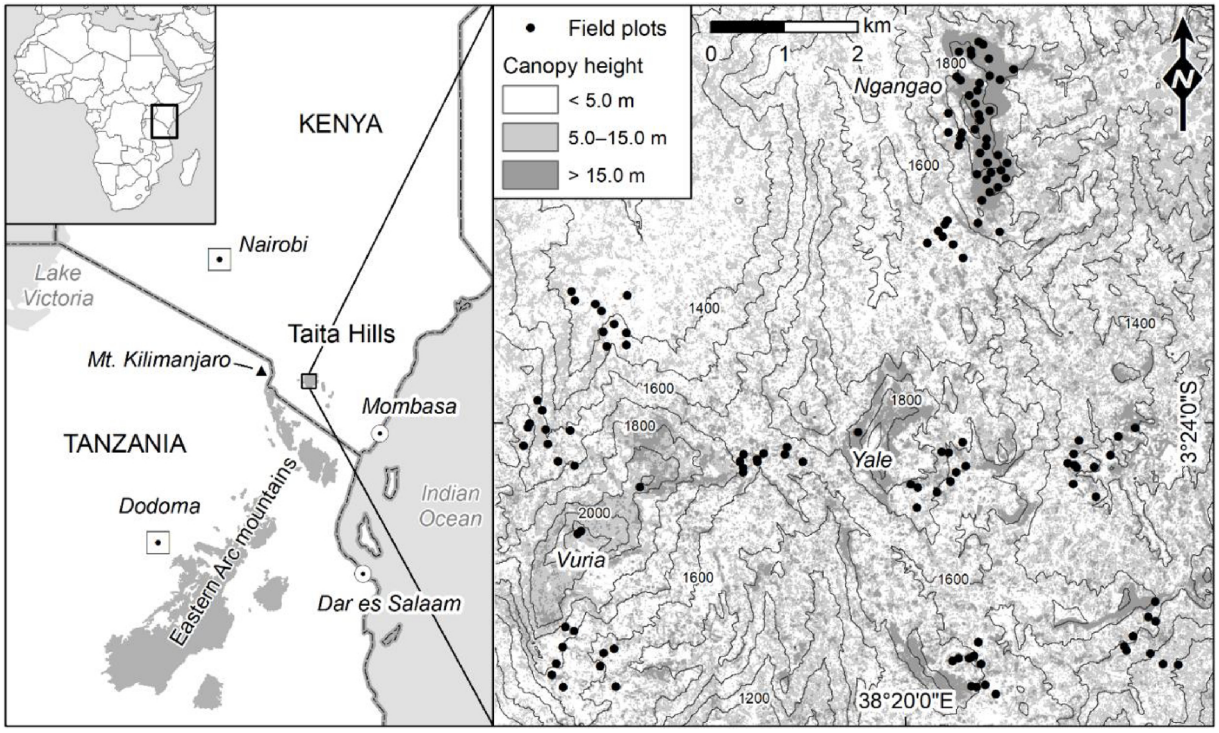

Fig. 1. Location of the study area and field plots. Canopy height model (CHM) and contour lines in 100 m interval are based on the airborne laser scanning data. CHM visualize the fine-scale variation in tree cover and height in the study area. The borders of the Eastern Arc mountains are from Platts et al. (2011).

Fig. 2. Typical landscape in the Taita Hills with settlements, maize fields and agroforestry, and montane forests and plantations in the mountain tops.

Table 1
Summary of the field plot data (n = 117).

Variable	Min	Max	Mean	SD
Stem density (stems ha^{-1})	10	1214	302	309
Mean diameter at breast height (cm)	13.2	45.5	23.1	6.5
Basal area (m^2 ha^{-1})	0.2	85.6	18.8	21.6
Basal area weighted mean height (m)	3.0	38.2	13.9	6.3
Maximum height (m)	3.0	48.0	19.3	8.7
Aboveground biomass (Mg ha^{-1})	0.3	642.2	120.7	160.5

Furthermore, we classified each plot as forest or non-forest in the field based on the dominant land cover. The classification was made according to the Land Degradation Surveillance Framework field guide (Vågen et al., 2013), which follows White (1983) vegetation type classification. The forest class included the montane forest plots, composed mainly of native tree species, and the plantation forest plots consisting mainly of exotics. The non-forest plots included everything else, such as croplands, agroforestry, and bushland. The number of the forest plots was 39 and the number of the non-forest plots was 78.

2.3. Airborne laser scanning data

We acquired ALS data 3–8 February 2013 by Optech ALTM 3100 sensor. The pulse rate was 100 kHz and scan rate 36 Hz. The maximum scan angle was ± 16 degrees, the mean range 760 m, and the mean footprint diameter 23 cm. A maximum of four returns were recorded per emitted pulse. As a result, the mean pulse density was 9.6 pulses m^{-2} and mean return density 11.4 returns m^{-2} (Pellikka et al., 2018). The acquisition period corresponds to the dry season when agricultural crops are or are soon going to be harvested. The most of the trees in the altitude of our study area do not shed their leaves during the dry season. The pre-processing of the ALS data was made by TopScan GmbH (Germany). Furthermore, we used Terrascan software (Terrasolid Oy, Finland) to filter the ground returns and buildings. We also filtered the powerlines and high points manually. The overlap of the neighboring flight lines was removed using lasoverage tool in LAStools software (rapidlasso GmbH, Germany). Finally, we used the ground returns to produce DEM at 1 m resolution.

2.4. Landsat time series

We used Landsat 8 Operational Land Imager (OLI) imagery for two full years of data since the first available Landsat 8 image for the study area (14 April 2013; path 167, row 62). The period corresponds roughly to the time of field data collection (2013–2014) although it begins about two months after the airborne data collection. However, we preferred to use higher quality Landsat 8 data instead of Landsat 7 Enhanced Thematic Mapper Plus data. We downloaded Landsat

Collection 1 Level-2 surface reflectance data from U.S. Geological Survey EarthExplorer[1] . The data included surface reflectance in blue (0.45–0.51 μm), green (0.53–0.59 μm), red (0.64–0.67 μm), near infrared, NIR (0.85–0.88 μm), and two shortwave infrared bands, SWIR1 (1.57–1.65 μm) and SWIR2 (2.11–2.29 μm), respectively. The spatial resolution of the OLI data is 30 m × 30 m. We masked clouds and cloud shadows using the CFMask cloud masks provided with the data (Foga et al., 2017). Furthermore, we performed topographic normalization using the C-correction method (Adhikari et al., 2016) and Japan Aerospace Exploration Agency (JAXA) DEM[2] .

Out of the 42 imagery available for the region, 30 were cloud and cloud shadow free in the 10 km × 10 km study area. In general, the cloudiness follows the variation of dry and wet seasons in the study area. The average percentage of cloud and cloud shadow cover was the smallest during the dry season in January (16.1%) and the largest during the rainy season in May (87%). The number of usable observations per pixel was 4–22 with median of 12 observations (6–17 with median of 11 observations in the field plots, respectively).

2.5. Hyperspectral data

We collected HS data simultaneously with ALS data using AisaEAGLE sensor (Specim, Spectral Imaging Ltd., Finland). The HS data had 129 bands in the spectral range of 400–1000 nm (bandwidth 4.5–5.0 nm). The planned flying height was approximately 750 m. The resulting pixel size was one meter (Piiroinen et al., 2017, 2018).

We made the radiometric correction and orthorectification for the raw data using CaliGeoPro 2.2 software (Specim, Spectral Imaging Ltd., Finland). We used a digital surface model derived from the ALS data in the orthorectification. After the correction, we noted that there were geometric mismatches between ALS and HS data. Therefore, we co-registered the HS data with the ALS data using control points collected from the CHM using second order polynomial transformation (Piiroinen et al., 2018). The atmospheric correction was done with ATCOR-4 software (Richter and Schläpfer, 2002). Furthermore, there were some cloud shadows in the southern part of the mosaic and we masked those manually.

Before calculating the predictor variables (Section 2.6), we also removed the darkest shadows and non-tree pixels from the mosaic as described by Piiroinen et al. (2018). The objective of this masking was to improve the sensitivity of the HS variables to the tree species composition. We masked all the pixels having NIR (836 nm) reflectance < 0.2 and normalized difference vegetation index (NDVI) (Rouse et al., 1973) < 0.5. These data were used for calculating variables for the HS only models (Section 2.7). However, as HS data alone does not allow reliable masking of the ground and lower vegetation pixels, we removed the remaining non-tree pixels from the data with the help of ALS-based CHM (height < 3 m). These data were used only for calculating variables for the models combining ALS and HS data (Section 2.7).

2.6. Predictor variables

We used ALS, LTS and HS data to compute a large number of predictor variables for AGB modelling (Supplement 1). In total, there were 66, 99 and 216 variables in ALS, LTS and HS data sets, respectively.

To calculate ALS variables, we extracted returns for the field plots using the radius of 17.84 m corresponding to the field plot radius and used FUSION software (McGaughey, 2018) for the variable computation. The computed variables were related to the laser return height distribution and canopy cover. We used a 3 m height threshold to separate canopy returns from ground and understory returns. This threshold was found earlier to work well with the 10 cm minimum DBH used in the field data collection (Adhikari et al., 2017; Pellikka et al., 2018). Furthermore, we computed the height-related variables separately using the first and the last returns only (Næsset, 2002). The first return data contained the 'single' and 'first of many' returns, while the last return data included the 'single' and 'last of many' returns.

For LTS data, we calculated a set of spectral-temporal variables for each pixel using all the cloud and cloud shadow free observations in the two years. In addition to the reflectance values in the spectral bands, we also calculated Tasseled Cap Brightness, Greenness and Wetness (Crist, 1985), NDVI (Rouse et al., 1973), and Reduced Simple Ratio (RSR) (Brown et al., 2000) vegetation indices. The Tasseled Cap indices were derived from the topographically corrected reflectance, and NDVI and RSR from the non-corrected reflectance as those are not affected by the topographic effects (Adhikari et al., 2016). The spectral-temporal variables included selected percentiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}) corresponding to the low, median, and high reflectance and vegetation index values in the time series. We also calculated trimmed means (10%, 25%) and inter-percentile ranges (10^{th}–90^{th}, 25^{th}–75^{th}) (Potapov et al., 2012; Hansen et al., 2016). The lowest and the highest values were excluded from mean and range due to possible unmasked clouds and cloud shadows. Finally, we used bilinear interpolation method to extract values for the field plot center locations.

HS variables were based on the pixels inside the field plots. The variables included spectral bands and narrowband VIs identified useful for the study of vegetation by Roberts et al. (2011) (see also Piiroinen et al., 2017). Furthermore, we applied the minimum noise fraction (MNF) transformation for data reduction (Green et al., 1988). The transformation was calculated using ENVI software (version 5.0, Research Systems Inc., USA). We used the first 15 MNF bands and disregarded the rest of the bands based on their low eigenvalues and visual analysis (Piiroinen et al., 2018). MNF transformation was calculated separately for two differently masked mosaics. For each variable, we calculated the trimmed mean (10%) and selected percentiles (10^{th}, 50^{th}, and 90^{th}) corresponding to the dark, median and bright pixels.

2.7. Predictive modelling

We applied linear regression for the AGB modelling. First, we tested ALS, LTS and HS variables separately, and then different combinations of the data sets. The models were also fit separately for the forest and non-forest plots.

We used 'lm' function in R Core Team (2016) for fitting the regression models. First, we searched the best models having 1–4 predictors by an exhaustive search using 'regsubsets' function in package 'leaps' (Lumley, 2017). For each data set and combination, we obtained the 30 best models. We applied a square root transformation for AGB as we found that it improved the model fits. When using models for prediction, we corrected the back-transformation bias by multiplying the predictions by the square of the standard error (Gregoire et al., 2008).

For accuracy assessment, we used leave one out cross-validation (LOOCV) (Packalén et al., 2012). For LOOCV, we removed each plot from the modelling set at the time and used it for validation. We repeated the procedure as many times as there were field plots, and computed accuracy statistics based on the observed and predicted values. The accuracy statistics included root mean square error (RMSE, Mg ha^{-1}) and relative RMSE (percentage of mean AGB), and pseudo coefficient of determination (R^2). We also combined the observed and LOOCV predicted values for the forest and non-forest strata in order to compare the non-stratified predictions and stratified predictions at the landscape-scale.

We applied several criteria to select the best model among the 120 model candidates for each set of plots and data. In order to assess if predicted values from LOOCV are consistent with observed values and the model is unbiased, we assessed if the predicted and observed values followed the 1:1 correspondence line using the hypothesis test (Piñeiro

[1] https://earthexplorer.usgs.gov/
[2] http://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm

Table 2
Modelling results for the different data sets. The non-stratified model included all the field plots. The forest and non-forest models were fit separately using forest and non-forest plots. The stratified models combine predictions of the best forest and non-forest models. Some of the ALS + LTS and ALS + HS + LTS combinations did not include variables from all the data sets and hence results area not shown. See Supplement 1 for the explanation of variable names.

Model	Data set	Selected variables	RMSE (Mg ha^{-1})	RMSE (%)	α	β	R^2	SSR	VIF_1	VIF_2	VIF_3	VIF_4
Non-stratified	ALS	hp25_f, cca	51.5	42.7	-2.5^{NS}	1.02^{NS}	0.90	1.04	3.3	3.3		
	LTS	RSR_p25, TCG_p50	85.1	70.5	-3.3^{NS}	1.11^{NS}	0.72	1.03	1.1	1.1		
	HS	PSRI_p10, mARI_p90, PRI_p90	89.8	74.4	-13.6^{NS}	1.10^{NS}	0.69	1.04	2.6	1.6	1.9	
	ALS + LTS	ALS variables only										
	ALS + HS	hp30_f, cca, MNF8_p90	48.9	40.5	-3.6^{NS}	1.03^{NS}	0.91	1.05	3.3	3.8	1.5	
	ALS + LTS + HS	ALS and HS variables only										
Forest	ALS	hp40_f, hsd_l	88.0	28.0	25.5^{NS}	0.92^{NS}	0.59	1.09	3.6	3.6		
	LTS	blue_q90, NDVI_range25_75, RSR_p90	118.0	37.5	34.8^{NS}	0.89^{NS}	0.26	1.09	1.6	1.3	1.2	
	HS	PSRI_p10, PRI_p90	123.0	39.1	62.8^{NS}	0.80^{NS}	0.21	1.08	1.3	1.3		
	ALS + LTS	hp25_f, NDVI_p25	82.3	26.2	11.4^{NS}	0.96^{NS}	0.64	1.08	1.1	1.1		
	ALS + HS	hp25_f, R722_p90, MNF8_p90	67.8	21.5	2.5^{NS}	0.99^{NS}	0.75	1.09	1.0	1.0	1.0	
	ALS + LTS + HS	ALS and HS variables only										
Non-forest	ALS	hmax_f, hL3_l, ccmeanaf	11.7	49.3	-1.2^{NS}	1.05^{NS}	0.82	1.07	2.3	1.6	1.9	
	LTS	blue_p90, SWIR1_p75	22.7	95.6	-6.5^{NS}	1.27^{NS}	0.35	1.03	3.8	3.8		
	HS	MNF13_p10, SIPI_p90, CRI1_p90	20.8	87.6	-6.0^{NS}	1.12^{NS}	0.45	1.07	1.5	1.8	1.3	
	ALS + LTS	hp30_f, ccmeanaf, blue_p90, SWIR1_p10	11.2	47.2	-0.4^{NS}	1.01^{NS}	0.84	1.08	1.5	2.1	2.7	3.2
	ALS + HS	hp60_l, ccaf, PSSR_p10, MNF12_p10	10.2	43.0	-0.8^{NS}	1.03^{NS}	0.86	1.09	1.4	1.8	1.4	1.0
	ALS + LTS + HS	ALS and HS variables only										
Stratified	ALS	ALS	51.7	42.8	1.7^{NS}	0.98^{NS}	0.90					
	LTS	LTS	70.7	58.6	1.0^{NS}	0.99^{NS}	0.81					
	HS	HS	72.7	60.2	2.1^{NS}	0.98^{NS}	0.79					
	ALS + LTS	ALS + LTS	48.4	40.1	0.7^{NS}	0.99^{NS}	0.91					
	ALS + HS	ALS + HS	40.0	33.1	0.1^{NS}	1.00^{NS}	0.94					
	ALS + LTS + HS	ALS and HS variables only										

ALS = airborne laser scanning; LTS = Landsat time series; HS = hyperspectral; RMSE = root mean square error based on leave-one-out cross validation; α/β = intercept/slope of observed versus predicted regression; ^{NS} = non-significant (null hypothesis not rejected, $p \geq 0.05$); SSR = sum of squares ratio; VIF_n = variance inflation factor for the nth selected variable in the model.

et al., 2008; Valbuena et al., 2017). The null hypothesis was that the intercept (α) and slope (β) of the linear regression between the predicted and observed values are 0 and 1, respectively. If the null hypotheses were rejected ($p < 0.05$), the model was not accepted (Valbuena et al., 2017). In order to evaluate the degree of overfitting to the sample, we employed the sum of squares ratio (SSR) (Valbuena et al., 2017). SSR is the ratio between the square root of the residual sums of squares (SS) from the cross-validation ($\sqrt{SS^{cv}}$) and without cross-validation ($\sqrt{SS^{fit}}$). SSR increase with overfitting, and we rejected models that had SSR > 1.1 because of possible overfitting (Valbuena et al., 2017). As many of the predictor variables were highly correlated, we also used variance inflation factor (VIF) to reject models with highly correlated variables (Zuur et al., 2010). We rejected the models having variables with VIF > 4. Furthermore, we accepted only models with significant predictor variables ($p < 0.05$). Finally, we selected the model with the lowest LOOCV RMSE among the models fulfilling all the above criteria.

3. Results

ALS data performed considerably better than LTS or HS data when used alone and without stratification (Table 2). LOOCV RMSE was 51.5 Mg ha^{-1} (42.7%) for ALS data with high pseudo R^2 (0.90). The predicted versus observed plots show that ALS model has some large random errors in the forests but in general the predictions follow 1:1 line across both strata (Fig. 3a). LTS model underestimated large AGB values, and both LTS and HS models overestimated small AGB values (Fig. 3b and c). The combination of ALS and HS data improved accuracy slightly (RMSE 48.9 Mg ha^{-1}, 40.5%) in comparison to the ALS only model (Fig. 3d). The best models combining ALS and LTS data did not contain any LTS variables (i.e. only ALS variables were selected).

ALS data were also clearly better than LTS and HS data for the forest and non-forest plots (Table 2). LTS and HS data performed similarly for the forest plots but HS data had particularly poor performance for the non-forest plots. However, the combination of ALS data with LTS and HS data improved accuracies of the forest and non-forest models. LOOCV RMSE of the forest models dropped from 88 Mg ha^{-1} (28.0%) to 82.3 Mg ha^{-1} (26.2%) and 67.8 Mg ha^{-1} (21.5%) when combining ALS data with LTS and HS data, respectively. Improvement of the non-forest models was not as considerable; LOOCV RMSE dropped from 11.7 Mg ha^{-1} (49.3%) to 11.2 Mg ha^{-1} (47.2%) and 10.2 Mg ha^{-1} (43.0%), respectively. Similar to the non-stratified models, the best models had only ALS and HS variables (i.e. no LTS variables were selected).

The accuracy of the ALS predictions at the landscape-level was not improved by using separate models for the forest and non-forest plots (Table 2, Fig. 3e). Nevertheless, the accuracy of the LTS and HS predictions was improved considerably in comparison to the non-stratified models, despite the poor model fits in terms of R^2 in the forest and non-forest plots (Table 2, Fig. 3f and g). Furthermore, the combined ALS and LTS data, and ALS and HS data, showed better accuracies than ALS model when the forest and non-forest strata were assessed together (Fig. 3h and i). Overall, the best accuracy at the landscape-level was achieved when using ALS and HS data with stratification (RMSE 40.0 Mg ha^{-1}, 33.1%).

The selected variables varied between the non-stratified and forest/non-forest models, and between the data combinations (Table 2). The best non-stratified ALS model included only two variables, one related to canopy cover and density (cca, percentage of all returns above 3 m) and another related to canopy height (hp25_f, 25th percentile of the first return heights). The best forest and non-forest models included other height related variables (hp30_f, hp40_f, hp60_l, hmax_f), standard deviation of the last return heights (hsd_l), L-moment 3 based on the last return heights (hL3_l), and other canopy cover and density related variables (ccmeanaf, ccaf). RSR based on red, NIR and SWIR1 reflectance was included in the non-stratified LTS model and forest

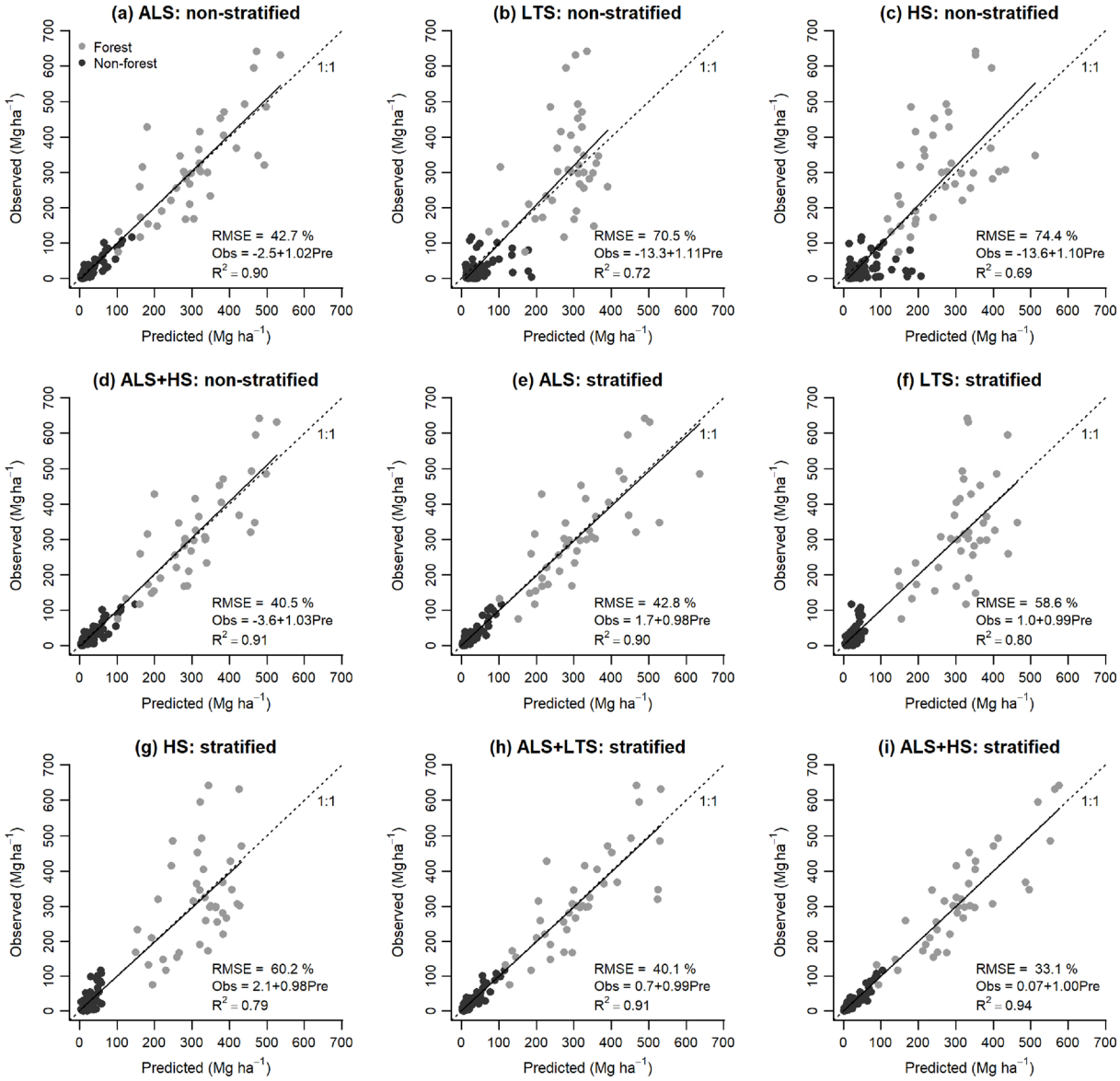

Fig. 3. Predicted versus observed aboveground biomass (Mg ha^{-1}) for the selected models based on leave-one-out cross validation. Non-stratified predictions (one model for all the field plots) based on (a) airborne laser scanning (ALS), (b) Landsat time series (LTS), (c) hyperspectral (HS) data, and (d) combined ALS and HS data. Stratified predictions (separate models for the forest and non-forest plots) based on (e) ALS, (f) LTS, (g) HS, (h) combined ALS and LTS, and (i) combined ALS and HS data.

model. However, the different percentile values corresponding to the dry season RSR (RSR_p25) and maximum RSR (RSR_p90) were used. Maximum reflectance in the blue band (blue_p90), corresponding to the dry season, was used in the forest and non-forest models. Other variables included median tasseled cap greenness (TCG_p50), NDVI interquartile range and 25th percentile (NDVI_range25_75, NDVI_p25), and 75th percentile reflectance in the SWIR1 band (SWIR1_p75). Furthermore, the best HS models included MNF bands and vegetation index (e.g. PSSR, PSRI and PRI). The model for forest included also the red edge spectral band at 722 nm. The selected HS variables were mostly low or high percentile values, i.e. values corresponding to the dark or bright pixels.

In order to apply the stratified models for prediction, it is important that those strata can be mapped with reasonable accuracy in order to apply separate models for them. In general, all the ALS variables used in the AGB prediction models can separate very well the forest and non-forest plots classified in the field (Fig. 4). However, only the number of returns above mean / total first returns × 100 (ccmeanaf) was able to separate forest and non-forest plots perfectly (Fig. 4d), and hence, we used it for predicting AGB (Fig. 5). The non-stratified ALS model, the model based on ALS and HS data, and the model based on ALS and HS model with stratification depict the same spatial patterns in AGB (Fig. 5a–c). The remaining forest patches are shown in the map as uniform areas of relatively high AGB, and abundant trees outside forests are visible in all the maps. However, as revealed by the difference images (Fig. 5d and e), large differences (> 10 Mg ha^{-1}) are observed locally, particularly in the forest areas where improvement in the accuracy was the greatest when combining ALS and HS data.

4. Discussion

We modelled AGB using ALS, LTS and HS data across an Afromontane landscape where trees are abundant outside forests. The most accurate predictions at the landscape-level were achieved by combining ALS and HS data, and using separate models for the forest and non-forest strata. In the forest strata, RMSE decreased from 28.0%

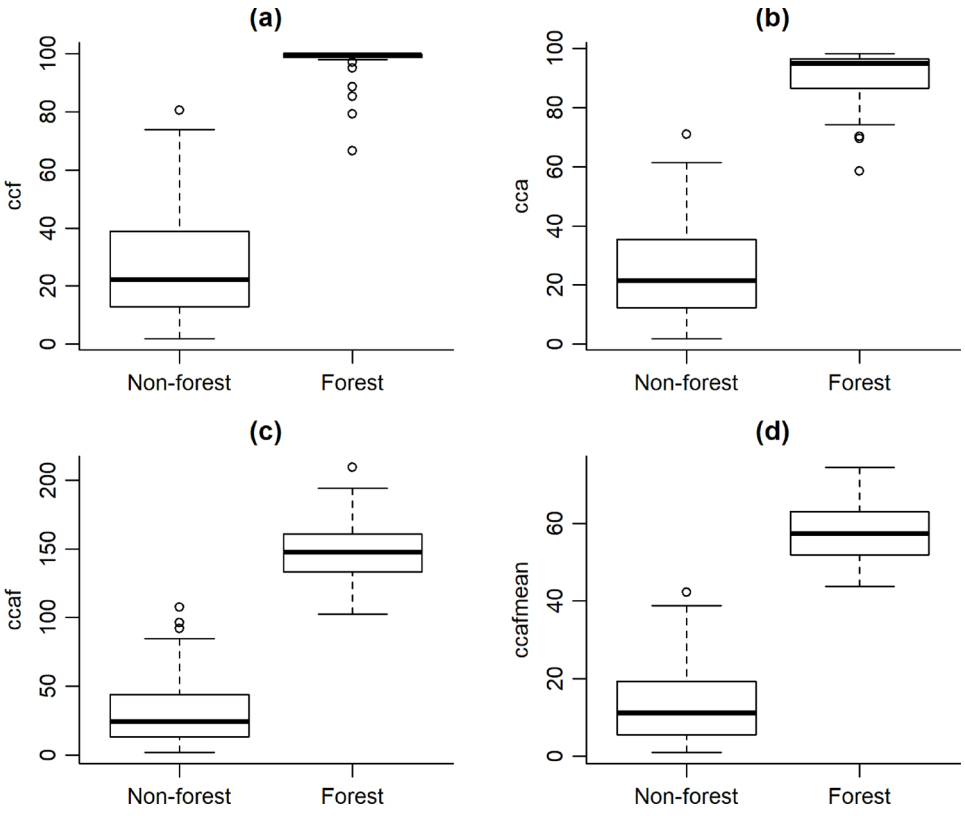

Fig. 4. Distribution of the selected canopy cover and density related airborne laser scanning variables for the non-forest and forest plots: (a) percentage of first returns above 3 m (ccf), (b) percentage of all returns above 3 m, (c) number of all returns above 3 m / total first returns × 100 (ccaf), and (d) number of returns above mean height / total first returns × 100 (ccmeanaf).

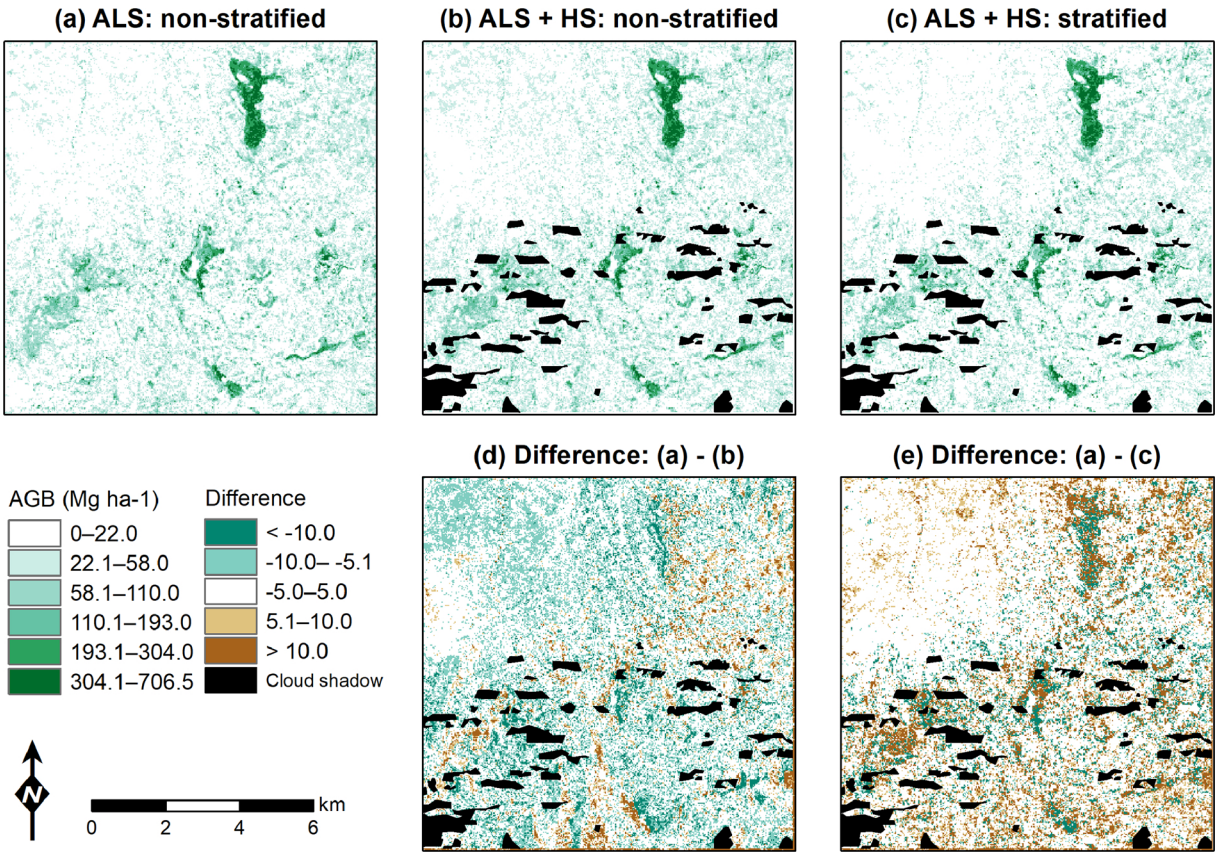

Fig. 5. Aboveground biomass (AGB) maps of the study area predicted by (a) the non-stratified model based on airborne laser scanning (ALS) data, (b) the non-stratified model based on ALS and hyperspectral (HS) data, and (c) the separate models for the forest and non-forest strata based on combined ALS and HS data. Panels (d) and (e) show differences between the maps (a) and (b), and (a) and (c), respectively. Part of the HS data were missing because of the cloud shadows.

to 21.5%, and in the non-forest strata from 49.3% to 43.0% in comparison to the ALS data. Both of these are major improvements and indicate possible synergies of the two data sets similar to Vaglio et al. (2014) and some studies outside Africa (Anderson et al., 2008). On the other hand, we observed minor improvement in the accuracy of the non-stratified predictions, which is in line with studies that have found only small benefits from additional HS data (Clark et al., 2011; Fassnacht et al., 2014). The variables in the best HS models included MNF bands and vegetation indices (e.g. pigment specific simple ratio, PSSR, Blackburn, 1998), which were used successfully for the tree species classification by Piiroinen et al. (2017). Therefore, HS data could improve AGB models through sensitivity to the tree species composition, which affect wood density. The selected HS variables can be sensitive to the season although most of the trees in the study area do not shed their leaves during the dry season. Therefore, the models combining ALS and HS data should be considered site and time specific.

Combining ALS and LTS resulted in minor improvements in the forest and non-forest plots. No improvement was observed in the case of non-stratified model. Earlier, Egberth et al. (2017) reported a small improvement in RMSE (from 50.3% to 49.1%) when combining ALS and Landsat 8 OLI data in the Miombo woodlands. In Australia, Ediriweera et al. (2014) found that combination of ALS and Landsat 5 data improved RMSE in the eucalyptus forest in but not in the closed-canopy subtropical forest. Therefore, small improvements in the prediction performance are possible when combining ALS and Landsat data. Also this improvement could be due to the sensitivity of the multispectral data on tree species composition. The selection of different percentile values and NDVI range emphasizes the potential of LTS for AGB prediction. However, because of the uneven distribution of the cloud-free observations over the year, the spectral-temporal variables are biased towards dry periods, which can limit their sensitivity to phenology. More data, for example from another sensor, would be necessary for improving observation distribution.

In general, our results demonstrate the strength of ALS data for AGB modelling across the tropical mosaic landscapes. LTS and HS data performed similarly with earlier studies using Landsat (Karlson et al., 2015; Gizachew et al., 2016; Egberth et al., 2017) and HS data (Fassnacht et al., 2014; Vaglio et al., 2014). Passive optical remote sensing data tends to saturate at relatively low AGB values, which was clearly observed also by the LTS data in this study. In contrast, ALS data are sensitive to the tree height and its variability, which makes it superior for modelling AGB (Zolkos et al., 2013). The cross-validated RMSE of the non-stratified ALS model (42.7%) is similar to the studies encompassing several vegetation types in Africa (Mauya et al., 2015; Næsset et al., 2016; Egberth et al., 2017).

The best non-stratified ALS model included only two variables; 25th height percentile based on the first returns, and canopy cover (or canopy density) based on all returns. Similar variables were used by Adhikari et al. (2017) and Pellikka et al. (2018). However, the forest and non-forest models included somewhat different variables (e.g., higher height percentiles and standard deviation of the last returns). This indicates that while some variables might be good predictors of AGB across a range of land cover types, others might be more suitable for the others. Typically, ALS-based AGB prediction models include both height and canopy cover related variables but otherwise, there seems to be little agreement on the exact predictor variables. Hansen et al. (2015) found that variables based on the returns from the lower parts of the canopy and canopy cover were more important than those describing the canopy height. On the other hand, Egberth et al. (2017) had 60th height percentile and two canopy cover variables based on the first returns in their model. Finding more general variables for the AGB mapping across Africa would require combining several AGB and ALS datasets from the different areas. Furthermore, sensitivity of ALS variables on season are poorly known in similar areas.

The selected ALS variables seem robust for a large range of the tree covered vegetation types. The forests in the Taita Hills show great variation in canopy structure (e.g., include species-rich tropical montane forest at different altitudes and single-species plantation forest) but have in general high canopy cover. On the other hand, the non-forest areas have sparse tree cover and sometimes the sample plots had only a few trees completely or partially in the plot. It could be expected that accuracy of the strata-specific ALS models is better than that of the general model as the strata are more homogenous (Mauya et al., 2015; Latifi et al., 2015). However, the forest and non-forest stratification was beneficial for the data fusion models but it did not improve the accuracy of the ALS-based predictions. This is similar to the results of Latifi et al. (2015). An earlier study by Heiskanen et al. (2015) suggested that leaf area index prediction models should be stratified according to the main forest types in the area. Hence, further studies should investigate stratification by the main forest types and by the dominant species of the plantations using a larger data set.

The stratification (stratified sampling or post-stratification) is sometimes considered impractical as it requires spatial data on vegetation or land cover, or some other relevant stratification, which is not necessarily trivial to produce (Mauya et al., 2015). Therefore, Mauya et al. (2015) recommended the non-stratified model when such data are not available. Here, we found that ALS variables, such as ccmeanaf, could separate forest and non-forest plots reliably, which enabled the application of the stratified models. In the case of LTS and HS only models, there was a clear improvement due to the post-stratification. However, the performance in terms of R^2 were poor for forest and non-forest classes in comparison to ALS models. Furthermore, using LTS or HS data alone, forest vs. non-forest classification would not have as good accuracy as attained by ALS data, also limiting the use of such models. In addition, the classification of the main forest types could be more difficult if stratification by the forest type or dominant species would be preferred.

ALS-based AGB modelling includes many uncertainties related to the field data collection. It is well-known that results improve when larger field plots are used (Zolkos et al., 2013). However, the field plots larger than 0.1 ha can be impractical to measure, particularly in the complex terrain and dense forests (Hansen et al., 2015). In the woodlands and croplands, larger plots would be more feasible as 0.1 ha plot include sometimes only a few trees. In these cases, large trees just outside the plot borders can have disproportionally large effect to the ALS variables although not included in AGB of the plot. Combination of the area-based and single-tree-based approaches could help reducing border effects due to the small plot size (Packalen et al., 2015). Uncertainties are also related to the allometric models, and tree height measurement and modelling (Chave et al., 2014; Valbuena et al., 2016), which increase modelling uncertainties. Furthermore, it has been common to measure only trees that have DBH $>$ 10 cm in the tropical forest inventories (Ganivet and Bloomberg, 2019). However, in some of the vegetation types (e.g., thickets and fallow fields), smaller DBH trees and shrubs could make a considerable share of the AGB, and affect the modelled relationships and selected variables. Therefore, further studies should pay more attention to the effect of the smaller trees and shrubs on the modelling accuracy.

5. Conclusions

The tropical mosaic landscapes have got increasing attention as those are vulnerable to land use and land cover change and large fraction of landscape-level AGB can reside outside forests. According to our results, ALS provides robust models for AGB prediction across such landscapes, including the areas outside forests. However, the results demonstrate a clear improvement in the modelling accuracy if combining ALS and HS data, and applying a forest and non-forest stratification. Furthermore, ALS data also provided the means for the forest and non-forest stratification, which is necessary for making use of HS data. Without stratification, only minor improvement in accuracy was achieved in comparison to using ALS data only. To conclude, our results

suggest that simultaneous ALS and HS data acquisition can be useful for the AGB modelling purposes in similar landscapes. The additional costs of HS data acquisition and processing can be also justified by other applications of HS data, such as tree species classification.

Acknowledgements

The data collection and work were supported by the Ministry for Foreign Affairs of Finland through the BIODEV, CHIESA and AFERIA projects. Petri Pellikka acknowledge funding from the Academy of Finland for the SMARTLAND project (decision number 318645). We thank Jessica Broas, Vuokko Heikinheimo, Elisa Schäfer, Jesse Hietanen and Darius Kizumi for assistance in the field and in data processing, and Taita Research Station of the University of Helsinki for logistics. Research permit from the National Council for Science and Technology of Kenya is gratefully acknowledged. Finally, we want to thank the two anonymous reviewers for their valuable comments.

References

Adhikari, H., Heiskanen, J., Maeda, E.E., Pellikka, P.K.E., 2016. The effect of topographic normalization on fractional tree cover mapping in tropical mountains: an assessment based on seasonal Landsat time series. Int. J. Appl. Earth Obs. Geoinf. 52, 20–31. https://doi.org/10.1016/j.jag.2016.05.008.

Adhikari, H., Heiskanen, J., Siljander, M., Maeda, E., Heikinheimo, V., Pellikka, P.K.E., 2017. Determinants of aboveground biomass across an afromontane landscape Mosaic in Kenya. Remote Sens. 9 (8), 827. https://doi.org/10.3390/rs9080827.

Albrecht, A., Kandji, S.T., 2003. Carbon sequestration in tropical agroforestry systems. Agric. Ecosyst. Environ. 99, 15–27. https://doi.org/10.1016/S0167-8809(03)00138-5.

Anderson, J.E., Plourde, L.C., Martin, M.E., Braswell, B.H., Smith, M., Dubayah, R.O., Hofton, M.A., Blair, J.B., 2008. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. Remote Sens. Environ. 112, 1856–1870. https://doi.org/10.1016/j.rse.2007.09.009.

Asner, G.P., Clark, J.K., Mascaro, J., Vaudry, R., Chadwick, K.D., Vieilledent, G., Rasamoelina, M., et al., 2012. Human and environmental controls over aboveground carbon storage in Madagascar. Carbon Balance Manag. 7, 2. https://doi.org/10.1186/1750-0680-7-2.

Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., et al., 2016. An integrated pan-tropical biomass map using multiple reference datasets. Glob. Change Biol. 22, 1406–1420. https://doi.org/10.1111/gcb.13139.

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., et al., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Change 2, 182–185. https://doi.org/10.1038/nclimate1354.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., Houghton, R.A., 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. Science 358, 230–234. https://doi.org/10.1126/science.aam5962.

Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. Int. J. Remote Sens. 19, 657–675. https://doi.org/10.1080/014311698215919.

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. Remote Sens. Environ. 206, 156–173. https://doi.org/10.1016/J.RSE.2017.12.030.

Brown, S., 1997. Estimating Biomass and Biomass Change of Tropical Forests: A Primer. FAO Forestry Paper 134.

Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. Remote Sens. Environ. 71, 16–25. https://doi.org/10.1016/S0034-4257(99)00035-8.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., et al., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Change Biol. 20, 3177–3190. https://doi.org/10.1111/gcb.12629.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., et al., 2013. Carbon and other biogeochemical cycles. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A. (Eds.), Climate Change 2013: The Physical Science Basis. Cambridge University Press, Cambridge, pp. 465–570.

Clark, M.L., Roberts, D.A., Ewel, J.J., Clark, D.B., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sens. Environ. 115, 2931–2942. https://doi.org/10.1016/j.rse.2010.08.029.

Crist, E.P., 1985. A TM tasseled cap equivalent transformation for reflectance factor data. Remote Sens. Environ. 17, 301–306. https://doi.org/10.1016/0034-4257(85)90102-6.

Curtis, R.O., 1967. Height-diameter and height-diameter-age equations for second growth Douglas fir. For. Sci. 13, 365–375.

Dube, T., Mutanga, O., 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. ISPRS J. Photogramm. Remote Sens. 101, 36–46. https://doi.org/10.1016/J.ISPRSJPRS.2014.11.001.

Ediriweera, S., Pathirana, S., Danaher, T., Nichols, D., 2014. Estimating above-ground biomass by fusion of LiDAR and multispectral data in subtropical woody plant communities in topographically complex terrain in North-eastern Australia. J. For. Res. 25, 761–771. https://doi.org/10.1007/s11676-014-0485-7.

Egberth, M., Nyberg, G., Næsset, E., Gobakken, T., Mauya, E., Malimbwi, R., Katani, J., et al., 2017. Combining airborne laser scanning and Landsat data for statistical modeling of soil carbon and tree biomass in Tanzanian Miombo woodlands. Carbon Balance Manag. 12, 8. https://doi.org/10.1186/s13021-017-0076-y.

Fassnacht, F.E., Hartig, F., Lati, H., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sens. Environ. 154, 102–114. https://doi.org/10.1016/j.rse.2014.07.028.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/J.RSE.2016.08.013.

Fatoyinbo, T., Feliciano, E.A., Lagomasino, D., Lee, S.K., Trettin, C., 2018. Estimating mangrove aboveground biomass from airborne LiDAR data: a case study from the Zambezi River delta. Environ. Res. Lett. 13, 025012. https://doi.org/10.1088/1748-9326/aa9f03.

Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Joseph Hughes, M., Laue, B., 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sens. Environ. 194, 379–390. https://doi.org/10.1016/J.RSE.2017.03.026.

Ganivet, E., Bloomberg, M., 2019. Towards rapid assessments of tree species diversity and structure in fragmented tropical forests: a review of perspectives offered by remotely-sensed and field-based data. For. Ecol. Manage. 432, 40–53. https://doi.org/10.1016/j.foreco.2018.09.003.

Gizachew, B., Solberg, S., Næsset, E., Gobakken, T., Bollandsås, O.M., Breidenbach, J., Zahabu, E., Mauya, E.W., 2016. Mapping and estimating the total living biomass and carbon in low-biomass woodlands using Landsat 8 CDR data. Carbon Balance Manag. 11, 13. https://doi.org/10.1186/s13021-016-0055-8.

Graves, S., Asner, G., Martin, R., Anderson, C., Colgan, M., Kalantari, L., Bohlman, S., et al., 2016. Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. Remote Sens. 8, 161. https://doi.org/10.3390/rs8020161.

Graves, S.J., Caughlin, T.T., Asner, G.P., Bohlman, S.A., 2018. A tree-based approach to biomass estimation from remote sensing data in a tropical agricultural landscape. Remote Sens. Environ. 218, 32–43. https://doi.org/10.1016/J.RSE.2018.09.009.

Green, A.A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Trans. Geosci. Remote Sens. 26, 65–74.

Gregoire, T.G., Lin, Q.F., Boudreau, J., Nelson, R., 2008. Regression estimation following the square root transformation of the response. For. Sci. 54, 597–606.

Hansen, E.H., Gobakken, T., Bollandsås, O.M., Zahabu, E., Næsset, E., 2015. Modeling aboveground biomass in dense tropical submontane rainforest using airborne laser scanner data. Remote Sens. 7, 788–807. https://doi.org/10.3390/rs70100788.

Hansen, M.C., Potapov, P.V., Goetz, S.J., Turubanova, S., Tyukavina, A., Krylov, A., Kommareddy, A., Egorov, A., 2016. Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. Remote Sens. Environ. 185, 221–232. https://doi.org/10.1016/j.rse.2016.02.023.

Heiskanen, J., Korhonen, L., Hietanen, J., Pellikka, P.K.E., 2015. Use of airborne lidar for estimating canopy gap fraction and leaf area index of tropical montane forests. Int. J. Remote Sens. 36, 2569–2583. https://doi.org/10.1080/01431161.2015.1041177.

Heiskanen, J., Liu, J., Valbuena, R., Aynekulu, E., Packalen, P., Pellikka, P., 2017. Remote sensing approach for spatial planning of land management interventions in West African savannas. J. Arid Environ. 140, 29–41. https://doi.org/10.1016/j.jaridenv.2016.12.006.

Henry, M., Picard, N., Trotta, C., Manlay, R., Valentini, R., Bernoux, M., Saint-André, L., 2011. Estimating tree biomass of sub-Saharan African forests: a review of available allometric equations. Silva Fenn. 45 (3B), 477–569. https://doi.org/10.14214/sf.38.

Houghton, R.A., Hall, F., Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. J. Geophys. Res. Biogeosci. 114, G00E03. https://doi.org/10.1029/2009JG000935.

ICRAF, 2015. Tree Functional Attributes and Ecological Database. World Agroforestry Centre (ICRAF) (Accessed 14 September 2018). http://db.worldagroforestry.org/.

IPCC, 2003. Good Practice Guidance for Land Use, Land Use Change and Forestry. IPCC National Greenhouse Gas Inventories Programme, Kanagawa, Japan.

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., Mattsson, E., Karlson, M., et al., 2015. Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest. Remote Sens. 7, 10017–10041. https://doi.org/10.3390/rs70810017.

Latifi, H., Fassnacht, F.E., Hartig, F., Berger, C., Hernández, J., Corvalán, P., Koch, B., 2015. Stratified aboveground forest biomass estimation by remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 38, 229–241. https://doi.org/10.1016/J.JAG.2015.01.016.

Laurance, W.F., Sayer, J., Cassman, K.G., 2014. Agricultural expansion and its impacts on tropical nature. Trends Ecol. Evol. 29, 107–116. https://doi.org/10.1016/j.tree.2013.12.001.

Lumley, T., 2017. leaps: Regression Subset Selection. R package version 3.0.

Marshall, A.R., Willcock, S., Platts, P.J., Lovett, J.C., Balmford, A., Burgess, N.D., Latham, J.E., et al., 2012. Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient. Biol. Conserv. 154, 20–33. https://doi.org/10.1016/J.BIOCON.2012.03.017.

Mauya, E.W., Ene, L.T., Bollandsås, O.M., Gobakken, T., Næsset, E., Malimbwi, R.E.,

Zahabu, E., 2015. Modelling aboveground forest biomass using airborne laser scanner data in the Miombo woodlands of Tanzania. Carbon Balance Manag. 10, 28. https://doi.org/10.1186/s13021-015-0037-2.
McGaughey, R.J., 2018. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. FUSION Version 3.80. United States Department of Agriculture, Forest Service, Pacific Northwest Research Station.
Mertz, O., Müller, D., Sikor, T., Hett, C., Heinimann, A., Castella, J.-C., Lestrelin, et al., 2012. The forgotten D: challenges of addressing forest degradation in complex mosaic landscapes under REDD+. Geogr. Tidsskr. J. Geogr. 112, 63–76. https://doi.org/10.1080/00167223.2012.709678.
Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.
Næsset, E., Ole, H., Solberg, S., Martin, O., Hofstad, E., Mauya, E., Zahabu, E., et al., 2016. Mapping and estimating forest area and aboveground biomass in Miombo woodlands in Tanzania using data from airborne laser scanning, TanDEM-X, RapidEye, and global forest maps: a comparison of estimated precision. Remote Sens. Environ. 175, 282–300. https://doi.org/10.1016/j.rse.2016.01.006.
Packalén, P., Temesgen, H., Maltamo, M., 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Can. J. Remote Sens. 38, 557–569. https://doi.org/10.5589/m12-046.
Packalen, P., Strunk, J.L., Pitkänen, J.A., Temesgen, H., Maltamo, M., 2015. Edge-tree correction for predicting forest inventory attributes using area-based approach with airborne laser scanning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 1274–1280. https://doi.org/10.1109/JSTARS.2015.2402693.
Paul, K.I., Roxburgh, S.H., England, J.R., Ritson, P., Hobbs, T., Brooksbank, K., John Raison, R., et al., 2013. Development and testing of allometric equations for estimating above-ground biomass of mixed-species environmental plantings. For. Ecol. Manage. 310, 483–494. https://doi.org/10.1016/J.FORECO.2013.08.054.
Pellikka, P.K.E., Lötjönen, M., Siljander, M., Lens, L., 2009. Airborne remote sensing of spatiotemporal change (1955–2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinf. 11, 221–232. https://doi.org/10.1016/J.JAG.2009.02.002.
Pellikka, P.K.E., Clark, B.J.F., Gosa, A.G., Himberg, N., Hurskainen, P., Maeda, E., Mwang'ombe, J., et al., 2013. Agricultural expansion and its consequences in the Taita Hills, Kenya. Dev. Earth Surf. Process. 16, 165–179. https://doi.org/10.1016/B978-0-444-59559-1.00013-X.
Pellikka, P.K.E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., Heiskanen, J., 2018. Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya. Appl. Geogr. 94, 178–189. https://doi.org/10.1016/j.apgeog.2018.03.017.
Phua, M.-H., Johari, S.A., Wong, O.C., Ioki, K., Mahali, M., Nilus, R., Coomes, D.A., et al., 2017. Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests. For. Ecol. Manag. 406, 163–171. https://doi.org/10.1016/J.FORECO.2017.10.007.
Piiroinen, R., Heiskanen, J., Mõttus, M., Pellikka, P., 2015. Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. Int. J. Appl. Earth Obs. Geoinf. 39, 1–8. https://doi.org/10.1016/j.jag.2015.02.005.
Piiroinen, R., Heiskanen, J., Maeda, E., Viinikka, A., Pellikka, P., 2017. Classification of tree species in a diverse African Agroforestry landscape using imaging spectroscopy and laser scanning. Remote Sens. 9, 875. https://doi.org/10.3390/rs9090875.
Piiroinen, R., Fassnacht, F.E., Heiskanen, J., Maeda, E., Mack, B., Pellikka, P., 2018. Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification. Remote Sens. Environ. 218, 119–131. https://doi.org/10.1016/J.RSE.2018.09.018.
Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol. Modell. 216, 316–322. https://doi.org/10.1016/J.ECOLMODEL.2008.05.006.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Core Team, R., 2015. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-121.
Platts, P.J., Burgess, N., Gereau, R., Lovett, J., Marshall, A., McClean, C., Pellikka, P., et al., 2011. Delimiting tropical mountain ecoregions for conservation. Environ. Conserv. 38, 312–324. https://doi.org/10.1017/S0376892911000191.
Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., Mane, L., Justice, C.O., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM + data. Remote Sens. Environ. 122, 106–116. https://doi.org/10.1016/J.RSE.2011.08.027.
R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. Int. J. Remote Sens. 23, 2631–2649. https://doi.org/10.1080/01431160110115834.
Roberts, D.A., Roth, K.L., Perroy, R.L., 2011. Hyperspectral Vegetation Indices. In: Thenkabail, P.S., Lyon, J.G., Huete, A. (Eds.), Thenkabail. Hyperspectral Remote Sensing of Vegetation. CRC press, Boca Raton, pp. 309–327.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great plains with ERTS. Third Earth Resources Technology Satellite-1 Symposium, National Aeronautics and Space Administration Vol. 1, 309–317.
Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., et al., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. U. S. A. 108, 9899–9904. https://doi.org/10.1073/pnas.1019576108.
Schnell, S., Kleinn, C., Ståhl, G., 2015. Monitoring trees outside forests: a review. Environ. Monit. Assess. 187, 600. https://doi.org/10.1007/s10661-015-4817-7.
Sloan, S., Sayer, J.A., 2015. Forest resources assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. For. Ecol. Manage. 352, 134–145. https://doi.org/10.1016/j.foreco.2015.06.013.
Tesfamichael, S.G., Beech, C., 2016. Combining Akaike's Information Criterion and discrete return LiDAR data to estimate structural attributes of savanna woody vegetation. J. Arid Environ. 129, 25–34. https://doi.org/10.1016/j.jaridenv.2016.02.006.
Vågen, T.-G., Winowiecki, L., Tamene Desta, L., Tondoh, J., 2013. The Land Degradation Surveillance Framework (LDSF) field Guide v3-2013. World Agroforestry Centre, Nairobi.
Vaglio, G., Chen, Q., Lindsell, J.A., Coomes, D.A., Del, F., Guerriero, L., Pirotti, F., Valentini, R., 2014. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. ISPRS J. Photogramm. Remote Sens. 89, 49–58. https://doi.org/10.1016/j.isprsjprs.2014.01.001.
Vaglio Laurin, G., Puletti, N., Chen, Q., Corona, P., Papale, D., Valentini, R., 2016. Above ground biomass and tree species richness estimation with airborne lidar in tropical Ghana forests. Int. J. Appl. Earth Obs. Geoinf. 52, 371–379. https://doi.org/10.1016/J.JAG.2016.07.008.
Valbuena, R., Heiskanen, J., Aynekulu, E., Pitkänen, S., Packalen, P., 2016. Sensitivity of above-ground biomass estimates to height-diameter modelling in mixed-species West African woodlands. PLoS One 11 (7), e0158198. https://doi.org/10.1371/journal.pone.0158198.
Valbuena, R., Hernando, A., Manzanera, J.A., Görgens, E.B., Almeida, D.R.A., Mauro, F., García-Abril, A., Coomes, D.A., 2017. Enhancing of accuracy assessment for forest above-ground biomass estimates obtained from remote sensing via hypothesis testing and overfitting evaluation. Ecol. Modell. 366, 15–26. https://doi.org/10.1016/J.ECOLMODEL.2017.10.009.
Vanderhaegen, K., Verbist, B., Hundera, K., Muys, B., 2015. REALU vs. REDD+: carbon and biodiversity in the afromontane landscapes of SW Ethiopia. For. Ecol. Manage. 343, 22–33. https://doi.org/10.1016/J.FORECO.2015.01.016.
White, F., 1983. The Vegetation of Africa. A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa. UNESCO, Paris.
Zanne, A.E., Lopez-Gonzalez, G., Coomes, D.A., Ilic, J., Jansen, S., Lewis, S.L., Miller, R.B., et al., 2009. Data from: Towards a worldwide wood economics spectrum. Dryad Digit. Repository. https://doi.org/10.5061/dryad.234.
Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS J. Photogramm. Remote Sens. 102, 222–231. https://doi.org/10.1016/J.ISPRSJPRS.2014.08.014.
Zolkos, S.G., Goetz, S.J., Dubayah, R., 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sens. Environ. 128, 289–298. https://doi.org/10.1016/j.rse.2012.10.017.
Zomer, R.J., Neufeldt, H., Xu, J., Ahrends, A., Bossio, D., Trabucco, A., van Noordwijk, M., Wang, M., 2016. Global Tree Cover and Biomass Carbon on Agricultural Land: the contribution of agroforestry to global and national carbon budgets. Sci. Rep. 6, 29987. https://doi.org/10.1038/srep29987.
Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods Ecol. Evol. 1, 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x.